

# Cumulative Distance Enumerators of Random Codes and their Thresholds

Yun Fan<sup>1</sup>, San Ling<sup>2</sup>, Hongwei Liu<sup>1</sup>, Jing Shen<sup>3</sup>, Chaoping Xing<sup>2</sup>

<sup>1</sup> School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China

<sup>2</sup> School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637616, Singapore

<sup>3</sup> College of Science, Naval University of Engineering, Wuhan, 430033, China

## Abstract

Cumulative weight enumerators of random linear codes are introduced, their asymptotic properties are studied, and very sharp thresholds are exhibited; as a consequence, it is shown that the asymptotic Gilbert-Varshamov bound is a very sharp threshold point for the density of the linear codes whose relative distance is greater than a given positive number. For arbitrary random codes, similar settings and results are exhibited; in particular, the very sharp threshold point for the density of the codes whose relative distance is greater than a given positive number is located at half the asymptotic Gilbert-Varshamov bound.

## 1 Introduction

Let  $F$  be an alphabet for coding with cardinality  $|F| = q > 1$ . Any non-empty subset  $C$  of  $F^n$  is called a code of length  $n$ ; the fraction  $R(C) = \frac{\log_q |C|}{n}$  is said to be the *rate* of the code  $C$ . Codewords, i.e. the elements of  $C$ , may be changed at some probability due to noise when they are communicated through a channel. Shannon [14] introduced the *capacity* of the channel and showed that: if a number  $r$  is below the capacity then there are codes of rate  $r$  and decoding rules such that the probability of recovering the codewords goes to 1 as  $n \rightarrow \infty$ ; however, if  $r$  is beyond the capacity then for any code of rate  $r$  the probability of recovering the codewords is bounded from above by a number less than 1. For the latter case, Wolfowitz [18] further showed that for any code of rate  $r$  the probability of recovering the codewords goes to 0 as  $n \rightarrow \infty$ . Following

---

*Email addresses:* yunfan02@yahoo.com.cn (Yun Fan); lingsan@ntu.edu.sg (San Ling); h\_w.liu@yahoo.com.cn (Hongwei Liu); shendina@hotmail.com (Jing Shen); xingcp@ntu.edu.sg (Chaoping Xing).

terminology from physics, Shannon's capacity is a *phase transition point*, also known as a *threshold point*, for the channel from high fidelity to no fidelity.

Hamming [6] introduced the *Hamming distance* to construct error-correcting codes. We denote the *minimum Hamming distance* of any code  $C \subseteq F^n$  by  $d_H(C)$ ; denote the fraction  $\frac{d_H(C)}{n}$  by  $\Delta(C)$ , and call it the *relative distance* of  $C$ . An important question is: how to get the highest rate of the codes in  $F^n$  with relative distance greater than a given positive number  $\delta$ ? Gilbert [5] and Varshamov [16] exhibited a function in  $\delta$  and  $n$ , called the *Gilbert-Varshamov bound*, to bound from below this highest rate, cf. [7, §2.8, §2.9]. The limit as  $n \rightarrow \infty$  of this lower bound, which we denote by  $GV_q(\delta)$ , is called the *asymptotic Gilbert-Varshamov bound*, or *GV-bound* in short. The GV-bound coincides exactly with Shannon's capacity of the  $q$ -ary symmetric channel, so it is also called the *Gilbert-Shannon-Varshamov bound*. When  $F$  is a finite field, there have been several families of linear codes in [15, 8, 17, 3] whose parameters improve the Gilbert-Varshamov bound (as a function in  $\delta$  and  $n$ ). In the asymptotic sense, the family in [15] does exceed the GV-bound but only for the case when  $q \geq 49$ .

For a random linear code  $L_r$  of rate  $r$ , Varshamov in [16] proved that: if  $r < GV_q(\delta)$ , then the probability of the relative distance of  $L_r$  being greater than  $\delta$ , denoted by  $\Pr(\Delta(L_r) > \delta)$ , goes to 1 as  $n \rightarrow \infty$ . Note that the relative distance  $\Delta(L_r)$  of the random linear code  $L_r$  of rate  $r$  is a random variable. Pierce [13] proved that  $\Delta(L_r)$  is asymptotically distributed at one point  $\delta_{GV}(r)$ , where  $\delta_{GV}(r)$  denotes the solution of the equation  $r = GV_q(x)$  in the variable  $x$  and is called the *GV-distance*. From this result of Pierce, it could further follow that: if  $r > GV_q(\delta)$  then  $\Pr(\Delta(L_r) > \delta)$  goes to 0 as  $n \rightarrow \infty$ . In other words, the GV-bound  $GV_q(\delta)$  is a threshold point from high density to extreme sparsity of the linear codes of rate  $r$  whose relative distance is greater than a given  $\delta$ .

For a binary random code  $C$  from a so-called *random code ensemble*, the *distance enumerator with respect to a given codeword*, i.e. the number of the codewords with a given distance from the given codeword, has been extensively studied; cf. [4], [9], [10]. Barg and Forney [1] investigated, by means of Chebyshev's inequality, the *distance enumerator* and showed that the relative distance  $\Delta(C)$  is asymptotically distributed at the point  $\delta_{GV}(2r)$ . Similar to the linear case but not as precisely as what Pierce [13] did, it could follow that  $\frac{1}{2}GV_q(\delta)$  is a threshold point for the density of the codes of rate  $r$  whose relative distance is greater than a given  $\delta$ .

In this paper we consider a more general setting with a more sensitive second moment method, including the Cauchy inequality and the Paley-Zygmund inequality (see [12]). For a random linear code  $L$  of length  $n$  and rate  $r$  we define the *cumulative weight enumerator* to be the number of codewords whose weight is bounded from above, and study its asymptotic property. We exhibit very sharp threshold points for the distributions of the cumulative weight enumerator. It is just a special case of the thresholds we obtain that the GV-bound  $GV_q(\delta)$  is a very sharp threshold point for the density of the linear codes of rate  $r$  whose relative distance is greater than a given  $\delta$ . Furthermore, the result

of [13] is also a special case of the threshold points for the cumulative weight enumerators.

For any random codes over an alphabet with  $q$  (not necessarily equal to 2) elements we consider a similar setting and obtain similar sharp threshold points. As immediate consequences, we show that  $\frac{1}{2}\text{GV}_q(\delta)$  is a very sharp threshold point for the density of the codes of rate  $r$  whose relative distance is greater than a given  $\delta$ ; on the other hand, we point out that the asymptotic distribution of the relative distances of the random codes with given rate  $r$  is exactly located at  $\delta_{\text{GV}}(2r)$ , as described in [1].

The paper is composed of five sections. After the introduction, in Section 2 the necessary notations are introduced for quotation. In Section 3 random linear codes are discussed; while in Section 4 arbitrary random codes are studied. Finally we draw conclusions in Section 5.

## 2 The KL-divergence and the GV-bound

Throughout the paper,  $F$  is an alphabet with cardinality  $|F| = q > 1$ ,  $n$  is a positive integer; and we always take the following parameters:

$$\delta \in (0, \delta_0) \text{ where } \delta_0 = 1 - q^{-1}, \quad d = \delta n; \quad r \in (0, 1), \quad k = rn; \quad (2.1)$$

where  $d = \delta n$  should be understood, in practical computations, to be the integer nearest to  $\delta n$ ; and it is the same for  $k = rn$ .

In this paper,  $\exp(x) = e^x$  denotes the natural exponential function with the natural base  $e$ , while  $\exp_q(x) = q^x$  denotes the exponential function with base  $q$ ; and  $\log_q x$  denotes the logarithm function to the base  $q$ , while  $\ln x$  denotes the natural logarithm function. Recall that, for any  $0 \leq p, p' \leq 1$ , the  $q$ -ary *KL-divergence* (or *information divergence*) of  $p$  to  $p'$  is defined as

$$D_q(p||p') = p \log_q \frac{p}{p'} + (1 - p) \log_q \frac{1-p}{1-p'}$$

with the conventions  $0 \log_q 0 = 0$  and  $0 \log_q \frac{0}{0} = 0$ . It is a convex non-negative function in  $p$  with a unique zero point at  $p = p'$ , see [11, §1.2].

We consider the following partial sum of binomials:

$$\beta_n(\delta) = \sum_{i=0}^{\delta n} \binom{n}{i} \delta_0^i (1 - \delta_0)^{n-i} = \sum_{i=0}^d \binom{n}{i} \delta_0^i (1 - \delta_0)^{n-i}, \quad (2.2)$$

where  $\binom{n}{i}$  denotes the binomial coefficient. We have the following estimation which will play a key role in the asymptotic analysis of random codes:

$$\beta_n(\delta) = \exp_q \left( -n D_q(\delta||\delta_0) - \frac{1}{2} \log_q n + \rho_n(\delta) \right) \quad (2.3)$$

with  $\rho_n(\delta)$ ,  $n = 1, 2, \dots$ , being functions on  $\delta$  such that

$$\lim_{n \rightarrow \infty} \exp_q \left( \rho_n(\delta) \right) = \frac{\sqrt{1-\delta}}{(1-\frac{\delta}{\delta_0})\sqrt{2\pi\delta}}.$$

We may also write

$$\beta_n(\delta) = \exp_q \left( -nD_q(\delta \parallel \delta_0) - \frac{1}{2} \log_q n + O(1) \right). \quad (2.3')$$

The estimation (2.3) is somewhat known, for example, [13, Formula (7)] is just a special case where  $q = 2$ . We sketch a proof of (2.3) for completeness.

**Proof of (2.3).** Let  $T_i = \binom{n}{i} \delta_0^i (1 - \delta_0)^{n-i}$ ; in particular,  $T_d = \binom{n}{\delta n} \delta_0^{\delta n} (1 - \delta_0)^{(1-\delta)n}$ . By Stirling's formula,  $n! = (2\pi n)^{\frac{1}{2}} n^n e^{-n+\gamma(n)}$  with  $\frac{1}{12n+1} < \gamma(n) < \frac{1}{12n}$ , so we have an estimation:

$$T_d = (2\pi\delta(1-\delta)n)^{-\frac{1}{2}} \exp_q \left( -nD_q(\delta \parallel \delta_0) - O\left(\frac{1}{n}\right) \right). \quad (2.4)$$

For any index  $i$  with  $0 < i < d$  we have

$$\frac{T_{i-1}}{T_i} = \frac{i}{n-i+1} \cdot \frac{1-\delta_0}{\delta_0} < \frac{i+1}{n-i} \cdot \frac{1-\delta_0}{\delta_0} = \frac{T_i}{T_{i+1}}.$$

For  $d = \delta n$  we let

$$\alpha = \frac{T_{d-1}}{T_d} = \frac{d(1-\delta_0)}{(n-d+1)\delta_0} = \frac{(\frac{1}{\delta_0}-1)\delta}{1-\delta+\frac{1}{n}}; \quad (2.5)$$

further, for  $\delta'$  with  $0 < \delta' < \delta$  and  $d' = \delta' n$ , we let

$$\alpha' = \frac{T_{d'-1}}{T_{d'}} = \frac{(\frac{1}{\delta'_0}-1)\delta'}{1-\delta'+\frac{1}{n}}.$$

Then

$$T_{d-i} = \frac{T_{d-i}}{T_{d-i+1}} \cdot \frac{T_{d-i+1}}{T_{d-i+2}} \cdots \frac{T_{d-1}}{T_d} \cdot T_d \leq \alpha^i T_d, \quad i = 0, 1, \dots, d;$$

and

$$T_{d-i} = \frac{T_{d-i}}{T_{d-i+1}} \cdot \frac{T_{d-i+1}}{T_{d-i+2}} \cdots \frac{T_{d-1}}{T_d} \cdot T_d \geq \alpha'^i T_d, \quad i = 0, 1, \dots, d-d'+1;$$

thus

$$\frac{1-\alpha'^{(\delta-\delta')n}}{1-\alpha'} T_d = T_d \sum_{i=0}^{d-d'-1} \alpha'^i \leq \sum_{i=0}^d T_i \leq T_d \sum_{i=0}^d \alpha^i = \frac{1-\alpha^{1+\delta n}}{1-\alpha} T_d;$$

that is,

$$\frac{(1-\alpha)(1-\alpha'^{(\delta-\delta')n})}{1-\alpha'} \leq \frac{\sum_{i=0}^d T_i}{(1-\alpha)^{-1} T_d} \leq 1 - \alpha^{1+\delta n}.$$

For any  $\varepsilon > 0$  we can take  $\delta' < \delta$  such that  $\frac{\alpha-\alpha'}{1-\alpha'} < \varepsilon$ ; hence  $\frac{1-\alpha}{1-\alpha'} > 1 - \varepsilon$ ; in this way we can get that

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=0}^d T_i}{(1-\alpha)^{-1} T_d} = 1. \quad (2.6)$$

Finally, combining (2.4), (2.5) and (2.6), we obtain the estimation (2.3).  $\square$

As an immediate consequence of (2.3'), we get the following lemma.

**Lemma 2.1.** *Let notations be as in (2.1) and (2.2). Then*

$$q^{rn}\beta_n(\delta) = \exp_q \left( (r - D_q(\delta\|\delta_0))n - \frac{1}{2} \log_q n + O(1) \right). \quad (2.7)$$

In particular,

$$\lim_{n \rightarrow \infty} q^{rn}\beta_n(\delta) = \begin{cases} 0, & r \leq D_q(\delta\|\delta_0); \\ \infty, & r > D_q(\delta\|\delta_0); \end{cases} \quad (2.8)$$

the limit is exponentially convergent for every  $r \in (0, 1)$  except for  $r = D_q(\delta\|\delta_0)$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.

Recall that any  $\mathbf{a} = (a_1, \dots, a_n) \in F^n$  is said to be a word of length  $n$  over  $F$ ; the *Hamming distance* between the words  $\mathbf{a}$  and  $\mathbf{a}'$ , denoted by  $d_H(\mathbf{a}, \mathbf{a}')$ , is defined to be the number of the indices  $i$  where  $\mathbf{a}$  and  $\mathbf{a}'$  have different coordinates  $a_i \neq a'_i$ . By  $B(\mathbf{a}, \delta n)$  we mean the Hamming ball centered at  $\mathbf{a} \in F^n$  with radius  $\delta n$ . The volume of the Hamming ball is denoted by  $V_q(n, \delta n) = |B(\mathbf{a}, \delta n)|$ . With the notation (2.2), we have

$$V_q(n, \delta n) = \sum_{i=0}^{\delta n} \binom{n}{i} (q-1)^i = q^n \beta_n(\delta). \quad (2.9)$$

Any non-empty subset  $C$  of  $F^n$  is called a  $q$ -ary code  $C$  of length  $n$  over  $F$ , any  $\mathbf{c} \in C$  is said to be a codeword. We can write the cardinality of  $C$  as  $|C| = q^{R(C)n}$ , where  $R(C) = \frac{\log_q(|C|)}{n}$  is called the *rate* of  $C$ . The minimum distance of  $C$ , denoted by  $d_H(C)$ , is defined as  $d_H(C) = \min_{\mathbf{c} \neq \mathbf{c}' \in C} d_H(\mathbf{c}, \mathbf{c}')$ ; and the fraction  $\frac{d_H(C)}{n}$  is called the *relative distance* of  $C$ , denoted by  $\Delta(C)$ .

For a given  $\delta \in (0, \delta_0)$ , a longstanding important question is the following: what is the highest rate  $R(C)$  for  $C \subseteq F^n$  with  $\Delta(C) > \delta$ ? We sketch it as a remark.

**Remark 2.1.** Applying a typical greedy algorithm to select codewords step by step, we can get a code  $C \subseteq F^n$  with  $\Delta(C) > \delta$  such that  $\bigcup_{\mathbf{c} \in C} B(\mathbf{c}, \delta n) = F^n$ , hence  $|C| \cdot V_q(n, \delta n) \geq q^n$ ; by Eqns (2.9) and (2.3') we have that

$$q^{R(C)n} \geq \frac{q^n}{V_q(n, \delta n)} = \exp_q \left( nD_q(\delta\|\delta_0) + \frac{1}{2} \log_q n - O(1) \right) \quad (2.10)$$

where  $\frac{q^n}{V_q(n, \delta n)}$  is just the usual *Gilbert-Varshamov bound* (cf. [7, §2.8, §2.9]), which is a function in  $n$  and  $\delta$ ; thus

$$R(C) \geq D_q(\delta\|\delta_0) + \frac{\log_q n}{2n} - O\left(\frac{1}{n}\right), \quad (2.11)$$

which implies that

- for large enough  $n$ , the highest rate of the codes of  $F^n$  whose relative distance is greater than  $\delta$  is larger than  $D_q(\delta\|\delta_0)$ .

A code  $C$  is said to be linear if  $F$  is a finite field and  $C$  is a subspace of the vector space  $F^n$ . For linear codes, applying a similar greedy algorithm to

construct *parity check matrices* of linear codes, we can get the same conclusions as above.

From the asymptotic perspective, the quantity  $D_q(\delta\|\delta_0)$  is known as the *asymptotic Gilbert-Varshamov bound* for  $q$ -ary codes at the relative distance  $\delta$ . We denote it by  $\text{GV}_q(\delta) = D_q(\delta\|\delta_0)$ , and call it the *GV-bound* in short. In fact, the GV-bound is also related to the so-called  *$q$ -ary entropy function*

$$H_q(\delta) = \delta \log_q(q-1) - \delta \log_q \delta - (1-\delta) \log_q(1-\delta)$$

in the sense that  $\text{GV}_q(\delta) = 1 - H_q(\delta)$ , see [7, §2.10.6].

As we have pointed out for the KL-divergence,  $\text{GV}_q(\delta)$  is a convex and strictly decreasing function in  $\delta \in (0, \delta_0)$  with  $\text{GV}_q(0) = 1$  and  $\text{GV}_q(\delta_0) = 0$ .

**Remark 2.2.** In the following, for convenience, we let  $r_\delta = \text{GV}_q(\delta)$  for any  $\delta \in (0, \delta_0)$ , and  $r_0 = 1$ ,  $r_{\delta_0} = 0$ ; i.e.  $r_\delta$  is the GV-bound at distance  $\delta$ . On the other hand, because of the monotone property of the function  $\text{GV}_q(\delta)$  in  $\delta \in (0, \delta_0)$ , for any  $r \in (0, 1)$  there is a unique real number in the interval  $(0, \delta_0)$ , which we denote by  $\delta_r$ , such that  $r = \text{GV}_q(\delta_r)$ ; of course,  $\delta_0 = 1 - q^{-1}$  as fixed in (2.1), and  $\delta_1 = 0$ . The quantity  $\delta_r$  is called the *GV-distance* at rate  $r$ ; see references [1], [10], [11, §6.2].

### 3 Random linear codes

In this section, we keep the parameters in (2.1) and assume that  $F$  is a finite field of order  $q$ , hence  $F^n$  is a vector space over  $F$  of dimension  $n$ . The Hamming weight  $w_H(\mathbf{a})$  for  $\mathbf{a} \in F^n$  is defined by  $w_H(\mathbf{a}) = d_H(\mathbf{a}, \mathbf{0})$ , where  $\mathbf{0}$  denotes the zero vector. Any linear subspace  $L$  of the vector space  $F^n$  is said to be a linear code; the rate of  $L$  is  $R(L) = \frac{\dim L}{n}$ , where  $\dim L$  denotes the dimension of the linear code  $L$ ; while the relative distance of  $L$  is  $\Delta(L) = \frac{w_H(L)}{n}$ , where  $w_H(L) = \min_{\mathbf{0} \neq \mathbf{c} \in L} w_H(\mathbf{c})$  is called the *minimum weight* of the linear code  $L$ .

Consider the probability space  $\mathcal{L}_r(F^n)$  whose sample space is the set of all linear codes of rate  $r$  in  $F^n$  and whose probability function is equiprobable. Take  $L_r \in \mathcal{L}_r(F^n)$ , i.e.  $L_r$  is a random linear code of length  $n$  and rate  $r$ . Set  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)$  to be the cardinality of the set of the non-zero codewords whose weight is at most  $\delta n$ , i.e.

$$\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) = |(L_r \setminus \{\mathbf{0}\}) \cap B(\mathbf{0}, \delta n)|, \quad (3.1)$$

where  $L_r \setminus \{\mathbf{0}\}$  denotes the set difference. In particular, the minimal  $\delta$  such that  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) \geq 1$  is just the relative distance  $\Delta(L_r)$  of  $L_r$ . We remark that  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) = \sum_{j=1}^{\delta n} \mathcal{N}_{L_r}^{\mathbf{0}}(j)$ , where  $\mathcal{N}_{L_r}^{\mathbf{0}}(j)$  stands for the number of codewords  $\mathbf{c}$  such that  $d_H(\mathbf{0}, \mathbf{c}) = j$ , and is called the *distance enumerator with respect to the codeword  $\mathbf{0}$* ; cf. [1, §B] or [11, §6.2]. We name  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)$  the *cumulative weight enumerator*.

For asymptotic analysis, we are concerned with the fraction  $\frac{\log_q \hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)}{n}$ , which is called the *growth rate* of  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)$ , and abbreviated by  $R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta))$ ; i.e.

$$\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) = \exp_q \left( n \cdot R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \right). \quad (3.2)$$

Note that  $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)$  and its growth rate are random variables over the probability space  $\mathcal{L}_r(F^n)$ . By Eqn (3.2),  $0 \leq \hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) \leq q^{rn} - 1$ , so

$$R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \in \{-\infty\} \cup [0, r). \quad (3.3)$$

Recalling from Remark 2.2 that  $r_\delta = \text{GV}_q(\delta)$  denotes the GV-bound at the distance  $\delta$ , we state the main result of this section.

**Theorem 3.1.** *Let notations be as in (3.2). Assume that  $\delta \in (0, \delta_0)$ ,  $r \in (0, 1)$  and  $t \in \{-\infty\} \cup [0, r)$ . Then*

$$\lim_{n \rightarrow \infty} \Pr \left( R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \geq t \right) = \begin{cases} 1, & \text{if } t < r - r_\delta; \\ 0, & \text{if } t \geq r - r_\delta; \end{cases}$$

*the limit converges exponentially for any  $t \in \{-\infty\} \cup [0, r)$  except for  $t = r - r_\delta$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

**Proof.** Let  $\mathcal{L}(F^k, F^n)$  be the probability space whose sample space is the set of all injective linear maps  $f : F^k \rightarrow F^n$  and every sample is equipped with equiprobability. For  $f \in \mathcal{L}(F^k, F^n)$  the image

$$L_f = f(F^k) = \{f(\mathbf{b}) \mid \mathbf{b} \in F^k\} \subseteq F^n \quad (3.4)$$

is a linear code of rate  $r$  in  $F^n$ , and any  $L_r \in \mathcal{L}_r(F^n)$  can be obtained in this way; further,  $L_f = L_{f'}$  for  $f, f' \in \mathcal{L}(F^k, F^n)$  if and only if  $f' = fg$  for a  $g \in \text{GL}(F^k)$  where  $\text{GL}(F^k)$  denotes the set of the linear automorphisms of  $F^k$ . In a word, the random linear code  $L_f$  and  $L_r$  have the same random property though they are defined in different probability spaces; in particular

$$\Pr \left( R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \geq t \right) = \Pr \left( R(\hat{\mathcal{N}}_{L_f}^{\mathbf{0}}(\delta)) \geq t \right).$$

Thus, instead of  $L_r$ , we prove the theorem for the random linear code  $L_f$  for  $f \in \mathcal{L}(F^k, F^n)$ .

First we write the random variable  $\hat{\mathcal{N}}_{L_f}^{\mathbf{0}}(\delta)$  as a sum of Bernoulli random variables. For each  $\mathbf{b} \in F^k \setminus \{\mathbf{0}\}$ , we have the following indicator random variable

$$X_{\mathbf{0}, \mathbf{b}} = \begin{cases} 1, & 0 < w_H(f(\mathbf{b})) \leq \delta n; \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

Since  $f(\mathbf{b})$  runs uniformly over the vectors of  $F^n \setminus \{\mathbf{0}\}$  when  $f$  runs uniformly over  $\mathcal{L}(F^k, F^n)$ , the expectation of  $X_{\mathbf{0}, \mathbf{b}}$  is easy to calculate (cf. Eqn (2.9)):

$$\mathbb{E}(X_{\mathbf{0}, \mathbf{b}}) = \Pr(X_{\mathbf{0}, \mathbf{b}} = 1) = \frac{V_q(n, \delta n) - 1}{q^n - 1} = \beta_n(\delta) - O(q^{-n}). \quad (3.6)$$

Further, we construct a non-negative integral random variable

$$X_{\mathbf{0}} = \sum_{\mathbf{b} \in F^k \setminus \{\mathbf{0}\}} X_{\mathbf{0}, \mathbf{b}}. \quad (3.7)$$

It is obvious that  $X_{\mathbf{0}} = \hat{\mathcal{N}}_{L_f}^{\mathbf{0}}(\delta) = \exp_q \left( n \cdot R(\hat{\mathcal{N}}_{L_f}^{\mathbf{0}}(\delta)) \right)$ , and  $R(\hat{\mathcal{N}}_{L_f}^{\mathbf{0}}(\delta)) \geq t$  is equivalent to  $X_{\mathbf{0}} \geq q^{tn}$ .

As the second step of the proof, we compute the expectation of the random variable  $X_{\mathbf{0}}$ . By the linearity of expectations and Eqn (3.6) we have

$$\begin{aligned} \mathbb{E}(X_{\mathbf{0}}) &= \sum_{\mathbf{b} \in F^k \setminus \{\mathbf{0}\}} \mathbb{E}(X_{\mathbf{0}, \mathbf{b}}) = (q^{rn} - 1)(\beta_n(\delta) - O(q^{-n})) \\ &= q^{rn} \beta_n(\delta)(1 - q^{-rn}) - q^{rn} O(q^{-n})(1 - q^{-rn}); \end{aligned}$$

thus, from Eqn (2.7), we obtain the following conclusion which we write as a lemma for later reference.

**Lemma 3.1.**  $\mathbb{E}(X_{\mathbf{0}}) = \exp_q \left( (r - r_{\delta})n - \frac{1}{2} \log_q n + O(1) \right) - O(q^{(r-1)n});$  in particular,  $\lim_{n \rightarrow \infty} \mathbb{E}(X_{\mathbf{0}}) = \begin{cases} 0, & r \leq r_{\delta}; \\ \infty, & r > r_{\delta}. \end{cases}$

The third step of the proof of Theorem 3.1 is proving the following lemma.

**Lemma 3.2.** If  $r > r_{\delta}$ , then  $\lim_{n \rightarrow \infty} \frac{\mathbb{E}(X_{\mathbf{0}})^2}{\mathbb{E}(X_{\mathbf{0}}^2)} = 1$ .

**Proof.** We fix a  $\mathbf{b}_1 \in F^k \setminus \{\mathbf{0}\}$ . For any  $\mathbf{b} \in F^k \setminus \{\mathbf{0}\}$  we have seen from Eqn (3.6) that  $\mathbb{E}(X_{\mathbf{0}, \mathbf{b}}) = \mathbb{E}(X_{\mathbf{0}, \mathbf{b}_1})$ ; and it is also obvious that

$$\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}} = 1) = \mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}_1} = 1),$$

where  $\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}} = 1)$  denotes the conditional expectation of  $X_{\mathbf{0}}$  when the event “ $X_{\mathbf{0}, \mathbf{b}} = 1$ ” occurs. By the lemma of [2, Appendix A] we have (or, one can check it directly):

$$\frac{\mathbb{E}(X_{\mathbf{0}})^2}{\mathbb{E}(X_{\mathbf{0}}^2)} = \sum_{\mathbf{b} \in F^k \setminus \{\mathbf{0}\}} \frac{\mathbb{E}(X_{\mathbf{0}, \mathbf{b}})}{\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}} = 1)} = \frac{(q^{rn} - 1) \cdot \mathbb{E}(X_{\mathbf{0}, \mathbf{b}_1})}{\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}_1} = 1)} = \frac{\mathbb{E}(X_{\mathbf{0}})}{\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}_1} = 1)}. \quad (3.8)$$

By the linearity of expectations, we have

$$\mathbb{E}(X_{\mathbf{0}} | X_{\mathbf{0}, \mathbf{b}_1} = 1) = \sum_{\mathbf{b} \in F^k \setminus \{\mathbf{0}\}} \mathbb{E}(X_{\mathbf{0}, \mathbf{b}} | X_{\mathbf{0}, \mathbf{b}_1} = 1).$$

For  $\mathbf{b} \in F^k \setminus \{\mathbf{0}\}$ , there are two cases.

*Case 1:*  $\mathbf{b}$  and  $\mathbf{b}_1$  are linearly independent. Since the event “ $X_{\mathbf{0}, \mathbf{b}_1} = 1$ ” occurs, we consider the injective linear maps  $f : F^k \rightarrow F^n$  which satisfy that  $f(\mathbf{b}_1) \in B(\mathbf{0}, \delta n)$ . When such  $f$  is randomly selected, the image  $f(\mathbf{b})$  runs over the vectors in  $F^n \setminus Ff(\mathbf{b}_1)$  uniformly at random, where  $Ff(\mathbf{b}_1)$  is the 1-dimensional subspace of  $F^n$  generated by  $f(\mathbf{b}_1)$ . Note that  $Ff(\mathbf{b}_1) \subseteq B(\mathbf{0}, \delta n)$



since any non-zero vector of  $Ff(\mathbf{b}_1)$  has weight  $w_H(f(\mathbf{b}_1))$ . Therefore, we get that

$$\begin{aligned} \mathbb{E}(X_{\mathbf{0},\mathbf{b}}|X_{\mathbf{0},\mathbf{b}_1} = 1) &= \Pr(X_{\mathbf{0},\mathbf{b}} = 1|X_{\mathbf{0},\mathbf{b}_1} = 1) = \frac{V_q(n,\delta n) - q}{q^n - q} \\ &< \frac{V_q(n,\delta n) - 1}{q^n - 1} = \mathbb{E}(X_{\mathbf{0},\mathbf{b}_1}) \end{aligned}$$

where the last equality follows from Eqn (3.6).

*Case 2:*  $\mathbf{b}$  and  $\mathbf{b}_1$  are linearly dependent. Then  $f(\mathbf{b})$  and  $f(\mathbf{b}_1)$  are linearly dependent, hence  $w_H(f(\mathbf{b})) = w_H(f(\mathbf{b}_1))$ ; so  $\mathbb{E}(X_{\mathbf{0},\mathbf{b}}|X_{\mathbf{0},\mathbf{b}_1} = 1) = 1$ .

Since the number of the non-zero  $\mathbf{b}$  which are linearly dependent on  $\mathbf{b}_1$  is equal to  $q - 1$ , from Eqn (3.8) we get that

$$\frac{\mathbb{E}(X_{\mathbf{0}}^2)}{\mathbb{E}(X_{\mathbf{0}})^2} = \frac{\sum_{\mathbf{b} \in F^k \setminus \{\mathbf{0}\}} \mathbb{E}(X_{\mathbf{0},\mathbf{b}}|X_{\mathbf{0},\mathbf{b}_1}=1)}{\mathbb{E}(X_{\mathbf{0}})} < \frac{(q-1) + (q^k - q)\mathbb{E}(X_{\mathbf{0},\mathbf{b}_1})}{\mathbb{E}(X_{\mathbf{0}})}.$$

Note that  $\frac{\mathbb{E}(X_{\mathbf{0}}^2)}{\mathbb{E}(X_{\mathbf{0}})^2} \geq 1$  (by Cauchy's inequality), and that  $(q^k - q)\mathbb{E}(X_{\mathbf{0},\mathbf{b}_1}) < (q^k - 1)\mathbb{E}(X_{\mathbf{0},\mathbf{b}_1}) = \mathbb{E}(X_{\mathbf{0}})$  (by (3.7)). We get that

$$1 \leq \frac{\mathbb{E}(X_{\mathbf{0}}^2)}{\mathbb{E}(X_{\mathbf{0}})^2} < \frac{q-1}{\mathbb{E}(X_{\mathbf{0}})} + 1.$$

Since  $r > r_\delta$ , by Lemma 3.1 we have that  $\lim_{n \rightarrow \infty} \mathbb{E}(X_{\mathbf{0}}) = \infty$ ; thus we get the desired conclusion  $\lim_{n \rightarrow \infty} \frac{\mathbb{E}(X_{\mathbf{0}})^2}{\mathbb{E}(X_{\mathbf{0}}^2)} = 1$ .  $\square$

Now we can complete the proof of Theorem 3.1.

Recall that  $\Pr(R(\hat{\mathcal{N}}_{L_G}^{\mathbf{0}}(\delta)) \geq t) = \Pr(X_{\mathbf{0}} \geq q^{tn})$ . If  $t \geq r - r_\delta$ , then, by Markov's inequality and Lemma 3.1, we have that

$$\Pr(X_{\mathbf{0}} \geq q^{tn}) \leq \frac{\mathbb{E}(X_{\mathbf{0}})}{q^{tn}} \leq \exp_q((r - r_\delta - t)n - \frac{1}{2} \log_q n + O(1));$$

since  $r - r_\delta - t \leq 0$ , we get  $\lim_{n \rightarrow \infty} \Pr(X_{\mathbf{0}} \geq q^{tn}) = 0$ .

In the case when  $t < r - r_\delta$ , we write  $q^{tn} = \theta \cdot \mathbb{E}(X_{\mathbf{0}})$  where

$$\frac{1}{\theta} = \frac{\mathbb{E}(X_{\mathbf{0}})}{q^{tn}} = \exp_q((r - r_\delta - t)n - \frac{1}{2} \log_q n + O(1)) - O(\exp_q((r - t - 1)n));$$

since  $r - r_\delta - t > 0$  and  $r - r_\delta - t > r - r_\delta - 1$ , we have  $\lim_{n \rightarrow \infty} \theta = 0$ . By the Paley-Zygmund inequality (see [12]) and Lemma 3.2, we get that

$$\lim_{n \rightarrow \infty} \Pr(X_{\mathbf{0}} \geq q^{tn}) = \lim_{n \rightarrow \infty} \Pr(X_{\mathbf{0}} \geq \theta \mathbb{E}(X_{\mathbf{0}})) \geq \lim_{n \rightarrow \infty} (1 - \theta)^2 \frac{\mathbb{E}(X_{\mathbf{0}})^2}{\mathbb{E}(X_{\mathbf{0}}^2)} = 1.$$

Finally, in all the above arguments (including the arguments for Lemma 3.2), the limits are exponentially convergent except for  $t = r - r_\delta$ ; at the exceptional point, the limit in Eqn (2.8) of Lemma 2.1 converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0. We are done for the proof of Theorem 3.1.  $\square$

Note that  $r-t \in (0, r] \cup \{\infty\}$  for  $t \in \{-\infty\} \cup [0, r)$ . Since  $\text{GV}_q(\delta)$  is a strictly decreasing function for  $\delta \in [0, \delta_0]$ , with the notations in Remark 2.2 we have:

$$t < r - r_\delta \iff r_\delta < r - t \iff \delta > \delta_{r-t};$$

$$t \geq r - r_\delta \iff r_\delta \geq r - t \iff \delta \leq \delta_{r-t};$$

where  $\delta_{r-t} \in [0, \delta_0)$  with the following convention:

$$\delta_{r-t} = 0 \quad \text{if} \quad r - t = \infty. \quad (3.9)$$

Thus we can rewrite Theorem 3.1 as follows.

**Theorem 3.1\*.** *Let notations be as in Theorem 3.1. With the convention in (3.9) we have that*

$$\lim_{n \rightarrow \infty} \Pr \left( R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \geq t \right) = \begin{cases} 0, & \text{if } \delta \leq \delta_{r-t}; \\ 1, & \text{if } \delta > \delta_{r-t}; \end{cases}$$

*the limit converges exponentially for any  $\delta \in (0, \delta_0)$  except for  $\delta = \delta_{r-t}$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

If we take  $t = 0$  and fix  $r$ , then the statement “ $R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta)) \geq 0$ ” is just saying that “ $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) \geq 1$ ”; and “ $\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta) \geq 1$ ” implies that the relative distance satisfies  $\Delta(L_r) \leq \delta$ . As a consequence of Theorem 3.1\* we obtain again the main result of [13] at once as follows.

**Corollary 3.1.** *Assume that  $r \in (0, 1)$  is given and  $\delta_r \in (0, \delta_0)$  is the corresponding GV-distance (see Remark 2.2). Then*

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta(L_r) \leq \delta \right) = \begin{cases} 0, & \delta \leq \delta_r; \\ 1, & \delta > \delta_r; \end{cases}$$

*the limit converges exponentially for any  $\delta \in (0, \delta_0)$  except for  $\delta = \delta_r$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

**Remark 3.1.** For any given  $r \in (0, 1)$ ,  $\Delta(L_r)$  is a random variable defined over the probability space  $\mathcal{L}_r(F^n)$ ; and  $\Pr(\Delta(L_r) \leq \delta)$  is just the cumulative distribution function of  $\Delta(L_r)$ . The above corollary shows that, asymptotically speaking, the distribution of  $\Delta(L_r)$  is concentrated at the GV-distance  $\delta_r$ ; more precisely, for arbitrarily small positive real number  $\varepsilon$  we have

$$\lim_{n \rightarrow \infty} \Pr(\Delta(L_r) \leq \delta_r) = 0 = \lim_{n \rightarrow \infty} \Pr(\Delta(L_r) \geq \delta_r + \varepsilon),$$

which is the main result of [13].

We now turn to Theorem 3.1. Recalling that the growth rate of the cumulative weight enumerator  $R(\hat{\mathcal{N}}_{L_r}^{\mathbf{0}}(\delta))$  is a random variable over  $\mathcal{L}_r(F^n)$ , we can

view Theorem 3.1 as a characterization of its distribution, i.e. for arbitrarily small  $\varepsilon > 0$  we have

$$\lim_{n \rightarrow \infty} \Pr(R(\hat{\mathcal{N}}_{L_r}^0(\delta)) \leq r - r_\delta - \varepsilon) = 0 = \lim_{n \rightarrow \infty} \Pr(R(\hat{\mathcal{N}}_{L_r}^0(\delta)) \geq r - r_\delta); \quad (3.10)$$

that is,  $R(\hat{\mathcal{N}}_{L_r}^0(\delta))$  is asymptotically distributed at exactly one point  $r - r_\delta$ . On the other hand, if we take  $t = 0$  again but fix  $\delta$ , then, since “ $\hat{\mathcal{N}}_{L_r}^0(\delta) < 1$ ” is equivalent to “ $\Delta(L_r) > \delta$ ”, we get a consequence from Theorem 3.1:

**Corollary 3.2.** *Assume that  $\delta \in (0, \delta_0)$  is given and  $r_\delta \in (0, 1)$  is the corresponding GV-bound (see Remark 2.2). Then*

$$\lim_{n \rightarrow \infty} \Pr(\Delta(L_r) > \delta) = \begin{cases} 1, & \text{if } r \leq r_\delta; \\ 0, & \text{if } r > r_\delta; \end{cases}$$

*the limit converges exponentially for any  $r \in (0, 1)$  except for  $r = r_\delta$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

The above corollary implies that the GV-bound  $r_\delta = \text{GV}_q(\delta)$  is a very sharp threshold point of the parameter  $r$  such that the density of the linear codes with relative distance greater than  $\delta$  changes from highly dense to extremely sparse.

**Remark 3.2.** More generally, for any point  $(\delta, r) \in [0, \delta_0] \times [0, 1]$  we can ask whether or not there exist linear codes  $L_r$  of rate  $r$  such that  $\Delta(L_r) > \delta$ . Our results show an asymptotically probabilistic answer as follows. The rectangle  $[0, \delta_0] \times [0, 1]$  can be divided into three areas as shown in Figure 3.1:

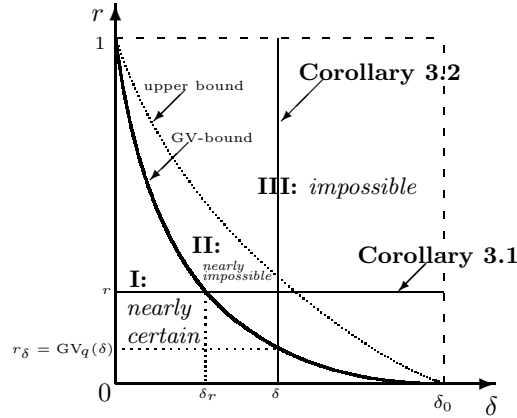


Figure 3.1. Three areas for the probability of the event “ $\Delta(L_r) > \delta$ ”

**Area I:** surrounded by the  $\delta$ -axis, the  $r$ -axis and the curve  $r = \text{GV}_q(\delta)$  (the curve itself included), where every point  $(\delta, r)$  satisfies  $\lim_{n \rightarrow \infty} \Pr(\Delta(L_r) > \delta) = 1$ ; in other words, “ $\Delta(L_r) > \delta$ ” is nearly a certain event. This fact is well known since Varshamov [16].

**Area II:** surrounded by the curve  $r = \text{GV}_q(\delta)$  and any upper bound (the dotted curve in Figure 3.1); for any point  $(\delta, r)$  in this area, we have  $\lim_{n \rightarrow \infty} \Pr(\Delta(L_r) > \delta) = 0$ ; that is, in the probabilistic sense, “ $\Delta(L_r) > \delta$ ” is nearly an impossible event. Note that, for a point  $(\delta, r)$  of this area, linear codes with rate  $r$  and with relative distance  $> \delta$  may exist, e.g. see the works [15, 8, 17, 3], see also Remark 2.1. However, in the asymptotic sense, the limits (as  $n \rightarrow \infty$ ) of the parameters of the linear codes in [8, 17, 3] are located on the curve  $r = \text{GV}_q(\delta)$ , hence in Area I, while the reference [15] does construct linear codes with parameters in Area II for the case where  $q \geq 49$ .

**Area III:** above any upper bound, where “ $\Delta(L_r) > \delta$ ” is an impossible event. There exist many upper bounds, for example, five upper bounds are described in [7, §2.10] (but the exact upper bound is still unknown). Our result also gives an explanation why it is so hard to determine the exact boundary of the area where “ $\Delta(L_r) > \delta$ ” is an impossible event: differently from the GV-bound where a very sharp phase transition occurs, the process from very small probability to zero probability is a very flat and calm stream, hence it seems hard to locate accurately the boundary from small to zero.

**Remark 3.3.** Instead of the injective linear maps from  $F^k$  to  $F^n$  as in Eqn (3.4), one can consider the probability space consisting of all the linear maps  $\tilde{f} : F^k \rightarrow F^n$  and with equiprobability. This probability space is just the so-called *random linear code ensemble* in Shannon’s information theory, cf. [11, §6.6]. Then, for the random linear codes  $\tilde{f}(F^k) \subseteq F^n$ , all the results of this section still hold.

## 4 Arbitrary random codes

In this section, the parameters in (2.1) are still preserved, but  $F$  is just an alphabet. Let  $C_r$  be a random code of rate  $r$  in  $F^n$ , in other words, we consider the probability space  $\mathcal{C}_r(F^n)$  consisting of all codes of rate  $r$  in  $F^n$  with equiprobability. The issues investigated in this section are similar to the last section, but the computations and the analysis of random dependence are different.

Similarly to Eqn (3.1), for  $\mathbf{c} \in C_r$  we define

$$\hat{\mathcal{N}}_{C_r}^{\mathbf{c}}(\delta) = |(C_r \setminus \{\mathbf{c}\}) \cap B(\mathbf{c}, \delta n)| = \sum_{j=1}^{\delta n} \mathcal{N}_{C_r}^{\mathbf{c}}(j), \quad (4.1)$$

where  $\mathcal{N}_{C_r}^{\mathbf{c}}(j)$  stands for the number of codewords  $\mathbf{c}'$  such that  $d_H(\mathbf{c}, \mathbf{c}') = j$ , and is called the *distance enumerator with respect to the codeword  $\mathbf{c}$* . Further, we set

$$\hat{\mathcal{N}}_{C_r}(\delta) = \frac{1}{2} \sum_{\mathbf{c} \in C_r} \hat{\mathcal{N}}_{C_r}^{\mathbf{c}}(\delta) = \frac{1}{2} \sum_{\mathbf{c} \in C_r} \sum_{j=1}^{\delta n} \mathcal{N}_{C_r}^{\mathbf{c}}(j), \quad (4.2)$$

which is the number of 2-subsets  $\{\mathbf{c}, \mathbf{c}'\}$  of  $C_r$  (i.e. subsets of  $C_r$  with cardinality 2) such that  $d_H(\mathbf{c}, \mathbf{c}') \leq \delta n$ . Thus, we name  $\hat{\mathcal{N}}_{C_r}(\delta)$  the *cumulative distance*

*enumerator.* In particular, the minimal  $\delta$  such that  $\hat{\mathcal{N}}_{C_r}(\delta) \geq 1$  is just the relative distance  $\Delta(C_r)$  of  $C_r$ . Similar to Eqn (3.2), we are concerned with the growth rate  $R(\hat{\mathcal{N}}_{C_r}(\delta)) = \frac{\log_q \hat{\mathcal{N}}_{C_r}(\delta)}{n}$ , i.e.

$$\hat{\mathcal{N}}_{C_r}(\delta) = \exp_q \left( n \cdot R(\hat{\mathcal{N}}_{C_r}(\delta)) \right). \quad (4.3)$$

Note that  $\hat{\mathcal{N}}_{C_r}(\delta)$  and its growth rate are random variables over the probability space  $\mathcal{C}_r(F^n)$ . From Eqn (4.2) we have that  $0 \leq \hat{\mathcal{N}}_{C_r}(\delta) \leq \frac{q^{rn}(q^{rn}-1)}{2}$ ; differently from Eqn (3.3), the range of  $R(\hat{\mathcal{N}}_{C_r}(\delta))$  is

$$R(\hat{\mathcal{N}}_{C_r}(\delta)) \in \{-\infty\} \cup [0, 2r). \quad (4.4)$$

**Theorem 4.1.** *Let notations be as in (4.3). Assume that  $\delta \in (0, \delta_0)$ ,  $r \in (0, 1)$  and  $t \in \{-\infty\} \cup [0, 2r)$ . Then*

$$\lim_{n \rightarrow \infty} \Pr \left( R(\hat{\mathcal{N}}_{C_r}(\delta)) \geq t \right) = \begin{cases} 1, & \text{if } t < 2r - r\delta; \\ 0, & \text{if } t \geq 2r - r\delta; \end{cases}$$

*the limit converges exponentially for any  $t \in \{-\infty\} \cup [0, 2r)$  except for  $t = 2r - r\delta$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

Similarly to the proof of Theorem 3.1, instead of  $C_r$ , we consider

$$C_g = \{g(\mathbf{b}) \mid \mathbf{b} \in F^k\} \subseteq F^n \quad (4.5)$$

for  $g \in \mathcal{C}(F^k, F^n)$ , where  $\mathcal{C}(F^k, F^n)$  is the probability space of all injective maps from  $F^k$  to  $F^n$  (recall that  $k = rn$ ) with equiprobability, since we have

$$\Pr \left( R(\hat{\mathcal{N}}_{C_r}(\delta)) \geq t \right) = \Pr \left( R(\hat{\mathcal{N}}_{C_g}(\delta)) \geq t \right).$$

We first reconstruct the integral random variable  $\hat{\mathcal{N}}_{C_g}(\delta)$  with Bernoulli variables. For any  $\mathbf{b} \neq \mathbf{b}' \in F^k$  we have the following indicator random variable

$$Y_{\mathbf{b}, \mathbf{b}'} = \begin{cases} 1, & d_H(g(\mathbf{b}), g(\mathbf{b}')) \leq \delta n; \\ 0, & \text{otherwise.} \end{cases} \quad (4.6)$$

Since  $g$  is injective, we get the expectation as follows:

$$\mathbb{E}(Y_{\mathbf{b}, \mathbf{b}'}) = \Pr(Y_{\mathbf{b}, \mathbf{b}'} = 1) = \frac{V_q(n, \delta n) - 1}{q^n - 1} = \beta_n(\delta) - O(q^{-n}). \quad (4.7)$$

**Remark 4.1.** For  $\mathbf{b} \in F^k$  we can define the integral random variable

$$Y_{\mathbf{b}} = \sum_{\mathbf{b}' \in F^k \setminus \{\mathbf{b}\}} Y_{\mathbf{b}, \mathbf{b}'} \quad (4.8)$$

so that  $Y_{\mathbf{b}} = \hat{\mathcal{N}}_{C_g}^{g(\mathbf{b})}(\delta)$ . Similar to what we did in Section 3, one can check that the conclusions of Lemma 3.1, Lemma 3.2 and Theorem 3.1 are all still true for  $Y_{\mathbf{b}}$ , i.e.

- (i).  $E(Y_{\mathbf{b}}) = \exp_q \left( (r - r_{\delta})n - \frac{1}{2} \log_q n - O(q^{(r-1)n}) \right);$
- (ii). if  $r > r_{\delta}$ , then  $\lim_{n \rightarrow \infty} \frac{E(Y_{\mathbf{b}})^2}{E(Y_{\mathbf{b}})} = 1;$
- (iii).  $\lim_{n \rightarrow \infty} \Pr \left( R(\hat{\mathcal{N}}_{C_g}^{g(\mathbf{b})}(\delta)) \geq q^{tn} \right) = \begin{cases} 1, & t < r - r_{\delta}; \\ 0, & t \geq r - r_{\delta}. \end{cases}$

However, in order to characterize  $\hat{\mathcal{N}}_{C_g}(\delta)$ , we need to introduce further an integral random variable:

$$Y = \frac{1}{2} \sum_{\mathbf{b} \in F^k} Y_{\mathbf{b}} = \sum_{\mathbf{b} \neq \mathbf{b}' \in F^k} Y_{\mathbf{b}, \mathbf{b}'}, \quad (4.9)$$

where  $\sum_{\mathbf{b} \neq \mathbf{b}' \in F^k}$  means that  $\{\mathbf{b}, \mathbf{b}'\}$  runs over the 2-subsets of  $F^k$ . Then we have that  $Y = \hat{\mathcal{N}}_{C_g}(\delta) = \exp_q \left( n \cdot R(\hat{\mathcal{N}}_{C_g}(\delta)) \right)$ , and the event “ $R(\hat{\mathcal{N}}_{C_g}(\delta)) \geq t$ ” is equivalent to the event “ $Y \geq q^{tn}$ ”.

**Lemma 4.1.**  $E(Y) = \exp_q \left( (2r - r_{\delta})n - \frac{1}{2} \log_q n + O(1) \right) - O(q^{(2r-1)n});$   
*in particular,*  $\lim_{n \rightarrow \infty} E(Y) = \begin{cases} 0, & 2r \leq r_{\delta}; \\ \infty, & 2r > r_{\delta}. \end{cases}$

**Proof.** Fix a 2-subset  $\{\mathbf{b}_1, \mathbf{b}_2\}$  of  $F^k$ . Let  $K = \frac{1}{2} q^{rn} (q^{rn} - 1)$ . Let  $\{\mathbf{b}, \mathbf{b}'\}$  be an arbitrary 2-subset of  $F^k$ . By (4.7) we have  $E(Y_{\mathbf{b}, \mathbf{b}'}) = E(Y_{\mathbf{b}_1, \mathbf{b}_2})$ ; further, by the linearity of expectations and Eqn (4.9), we have

$$E(Y) = K \cdot E(Y_{\mathbf{b}_1, \mathbf{b}_2}) = K \cdot (\beta_n(\delta) - O(q^{-n})).$$

As  $K\beta_n(\delta) = \frac{1}{2} q^{2rn} \beta_n(\delta) (1 - q^{-rn})$ , the first equality of the lemma follows from Lemma 2.1. Then, it is clear that  $\lim_{n \rightarrow \infty} E(Y) = 0$  if  $2r \leq r_{\delta}$ ; on the other hand, since  $2r - r_{\delta} > 2r - 1$ , we have that  $\lim_{n \rightarrow \infty} E(Y) = \infty$  if  $2r > r_{\delta}$ .  $\square$

**Lemma 4.2.** *If  $2r > r_{\delta}$ , then  $\lim_{n \rightarrow \infty} \frac{E(Y)^2}{E(Y^2)} = 1$ .*

**Proof.** Keep the notations as in the proof of Lemma 4.1. Besides the equality  $E(Y_{\mathbf{b}, \mathbf{b}'}) = E(Y_{\mathbf{b}_1, \mathbf{b}_2})$ , it is also obvious that  $E(Y|Y_{\mathbf{b}, \mathbf{b}'} = 1) = E(Y|Y_{\mathbf{b}_1, \mathbf{b}_2} = 1)$ . Similar to the proof of Lemma 3.2, we have:

$$\frac{E(Y)^2}{E(Y^2)} = \sum_{\mathbf{b} \neq \mathbf{b}' \in F^k} \frac{E(Y_{\mathbf{b}, \mathbf{b}'})}{E(Y|Y_{\mathbf{b}, \mathbf{b}'} = 1)} = \frac{K \cdot E(Y_{\mathbf{b}_1, \mathbf{b}_2})}{E(Y|Y_{\mathbf{b}_1, \mathbf{b}_2} = 1)} = \frac{E(Y)}{E(Y|Y_{\mathbf{b}_1, \mathbf{b}_2} = 1)}, \quad (4.10)$$

and

$$E(Y|Y_{\mathbf{b}_1, \mathbf{b}_2} = 1) = \sum_{\mathbf{b} \neq \mathbf{b}' \in F^k} E(Y_{\mathbf{b}, \mathbf{b}'} | Y_{\mathbf{b}_1, \mathbf{b}_2} = 1).$$

For an arbitrary 2-subset  $\{\mathbf{b}, \mathbf{b}'\}$  of  $F^k$ , there are three cases.

*Case 1:*  $\{\mathbf{b}, \mathbf{b}'\} = \{\mathbf{b}_1, \mathbf{b}_2\}$ , then  $E(Y_{\mathbf{b}, \mathbf{b}'} | Y_{\mathbf{b}_1, \mathbf{b}_2} = 1) = 1$ .

*Case 2:*  $\{\mathbf{b}, \mathbf{b}'\}$  and  $\{\mathbf{b}_1, \mathbf{b}_2\}$  have one vector in common, the number of such 2-subsets is  $2(q^{rn} - 2)$ ; for this case, we have

$$E(Y_{\mathbf{b}, \mathbf{b}'} | Y_{\mathbf{b}_1, \mathbf{b}_2} = 1) = \frac{V_q(n, \delta n) - 2}{q^n - 2} \leq \frac{V_q(n, \delta n) - 1}{q^n - 1} = E(Y_{\mathbf{b}_1, \mathbf{b}_2}).$$

*Case 3:*  $\{\mathbf{b}, \mathbf{b}'\} \cap \{\mathbf{b}_1, \mathbf{b}_2\} = \emptyset$ . Since  $g$  is injective, we can bound the conditional expectation as follows

$$E(Y_{\mathbf{b}, \mathbf{b}'} | Y_{\mathbf{b}_1, \mathbf{b}_2} = 1) \leq \frac{V_q(n, \delta n) - 3}{q^n - 3} \leq \frac{V_q(n, \delta n) - 1}{q^n - 1} = E(Y_{\mathbf{b}_1, \mathbf{b}_2}).$$

By Eqn (4.10) we have (recall that  $K \cdot E(Y_{\mathbf{b}_1, \mathbf{b}_2}) = E(Y)$ ):

$$1 \leq \frac{E(Y^2)}{E(Y)^2} = \frac{E(Y | Y_{\mathbf{b}_1, \mathbf{b}_2} = 1)}{E(Y)} \leq \frac{1 + (K-1)E(Y_{\mathbf{b}_1, \mathbf{b}_2})}{E(Y)} < \frac{1}{E(Y)} + 1.$$

Because  $2r > r_\delta$ , by Lemma 4.1 we have  $\lim_{n \rightarrow \infty} E(Y) = \infty$ . Thus we get the desired conclusion  $\lim_{n \rightarrow \infty} \frac{E(Y^2)}{E(Y)^2} = 1$ .  $\square$

Recalling that “ $R(\hat{\mathcal{N}}_{C_g}(\delta)) \geq t$ ” is equivalent to “ $Y \geq q^{tn}$ ” (see Eqn (4.9)), we prove Theorem 4.1 now.

**Proof of Theorem 4.1.** First assume that  $t \geq 2r - r_\delta$ . By Markov’s inequality and Lemma 4.1, we have that

$$\Pr(Y \geq q^{tn}) \leq \frac{E(Y)}{q^{tn}} \leq \exp_q((2r - t - r_\delta)n - \frac{1}{2} \log_q n + O(1));$$

since  $2r - t - r_\delta \leq 0$ , we get  $\lim_{n \rightarrow \infty} \Pr(Y \geq q^{tn}) = 0$ .

Next, we assume that  $t < 2r - r_\delta$ . We write  $q^{tn} = \theta \cdot E(Y)$  where

$$\frac{1}{\theta} = \frac{E(Y)}{q^{tn}} = \exp_q((2r - t - r_\delta)n - \frac{1}{2} \log_q n + O(1)) - O(\exp_q((2r - t - 1)n));$$

since  $2r - t - r_\delta > 2r - t - 1$ , we have that  $\lim_{n \rightarrow \infty} \theta = 0$ ; by the Paley-Zygmund inequality (see [12]) and Lemma 4.2, we get that

$$\lim_{n \rightarrow \infty} \Pr(Y \geq q^{tn}) = \lim_{n \rightarrow \infty} \Pr(Y \geq \theta E(Y)) \geq \lim_{n \rightarrow \infty} (1 - \theta)^2 \frac{E(Y)^2}{E(Y^2)} = 1.$$

Finally, in all the arguments above (including the arguments for Lemma 4.2), the limits are exponentially convergent except for  $t = 2r - r_\delta$ ; at the exceptional point, the limit in (2.8) of Lemma 2.1 converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0. We are done for the proof.  $\square$

Note that  $2r - t \in (0, 2r] \cup \{\infty\}$  for  $t \in \{-\infty\} \cup [0, 2r)$ . Since  $\text{GV}_q(\delta)$  is a strictly decreasing function for  $\delta \in [0, \delta_0]$ , we have

$$t < 2r - r_\delta \iff r_\delta < 2r - t \iff \delta > \delta_{2r-t};$$

$$t \geq 2r - r_\delta \iff r_\delta \geq 2r - t \iff \delta \leq \delta_{2r-t};$$

where the GV-distance  $\delta_{2r-t} \in [0, \delta_0)$  with the following convention:

$$\delta_{2r-t} = 0 \quad \text{if} \quad 1 < 2r - t. \quad (4.11)$$

Similar to Theorem 3.1\*, we can rewrite Theorem 4.1 as follows.

**Theorem 4.1\*.** *Let notations be as in Theorem 4.1. With the convention in (4.11) we have that*

$$\lim_{n \rightarrow \infty} \Pr \left( R(\hat{\mathcal{N}}_{C_r}(\delta)) \geq t \right) = \begin{cases} 0, & \text{if } \delta \leq \delta_{2r-t}; \\ 1, & \text{if } \delta > \delta_{2r-t}; \end{cases}$$

*the limit converges exponentially for any  $\delta \in (0, \delta_0)$  except for  $\delta = \delta_{2r-t}$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

If we take  $t = 0$  and fix  $r$ , then the statement “ $R(\hat{\mathcal{N}}_{C_r}(\delta)) \geq 0$ ” is none other than “ $\hat{\mathcal{N}}_{C_r}(\delta) \geq 1$ ”; and “ $\hat{\mathcal{N}}_{C_r}(\delta) \geq 1$ ” is equivalent to “ $\Delta(C_r) \leq \delta$ ”, we get a straightforward consequence of Theorem 4.1\* as follows.

**Corollary 4.1.** *Assume that  $r \in (0, 1)$  is given and  $\delta_{2r} \in [0, \delta_0)$  is the GV-distance at the rate  $2r$  with the convention in (4.11). Then*

$$\lim_{n \rightarrow \infty} \Pr \left( \Delta(C_r) \leq \delta \right) = \begin{cases} 0, & \delta \leq \delta_{2r}; \\ 1, & \delta > \delta_{2r}; \end{cases}$$

*the limit converges exponentially for any  $\delta \in (0, \delta_0)$  except for  $\delta = \delta_{2r}$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

**Remark 4.2.** Similar to the explanation in Remark 3.1,  $\Delta(C_r)$  is a random variable over the probability space  $\mathcal{C}_r(F^n)$ , and  $\Pr(\Delta(C_r) \leq \delta)$  is its cumulative distribution function. The corollary exhibits that the distribution of  $\Delta(C_r)$  is asymptotically concentrated at the GV-distance  $\delta_{2r}$ ; more precisely, for an arbitrarily small positive real number  $\varepsilon$  we have

$$\lim_{n \rightarrow \infty} \Pr(\Delta(C_r) \leq \delta_{2r}) = 0 = \lim_{n \rightarrow \infty} \Pr(\Delta(C_r) \geq \delta_{2r} + \varepsilon).$$

The references [1], [10] have also described the distribution, but in terms of the distance enumerator with respect to any given codeword, and not as precisely as above.

We turn to Theorem 4.1. Similar to Eqn (3.10),  $R(\hat{\mathcal{N}}_{C_r}(\delta))$  is a random variable over  $\mathcal{C}_r(F^n)$ , and for an arbitrarily small  $\varepsilon > 0$  we have

$$\lim_{n \rightarrow \infty} \Pr(R(\hat{\mathcal{N}}_{C_r}(\delta)) \leq 2r - r_\delta - \varepsilon) = 0 = \lim_{n \rightarrow \infty} \Pr(R(\hat{\mathcal{N}}_{C_r}(\delta)) \geq 2r - r_\delta); \quad (4.12)$$



that is,  $R(\hat{\mathcal{N}}_{C_r}(\delta))$  is asymptotically distributed at exactly one point  $2r - r_\delta$ . On the other hand, if we take  $t = 0$  again but fix  $\delta$ , then, since “ $\hat{\mathcal{N}}_{C_r}(\delta) = 0$  (i.e.  $\hat{\mathcal{N}}_{C_r}(\delta) < 1$ )” is equivalent to “ $\Delta(C_r) > \delta$ ”, we get a corollary as follows.

**Corollary 4.2.** *Assume that  $\delta \in (0, \delta_0)$  is given and  $r_\delta \in (0, 1)$  is the corresponding GV-bound. Then*

$$\lim_{n \rightarrow \infty} \Pr(\Delta(C_r) > \delta) = \begin{cases} 1, & \text{if } r \leq \frac{1}{2}r_\delta; \\ 0, & \text{if } r > \frac{1}{2}r_\delta; \end{cases}$$

*the limit converges exponentially for any  $r \in (0, 1)$  except for  $r = \frac{1}{2}r_\delta$ ; at the exceptional point it converges to 0 at the rate of  $\frac{1}{\sqrt{n}}$  going to 0.*

**Remark 4.3.** The above corollary says that  $\frac{1}{2}r_\delta = \frac{1}{2}\text{GV}_q(\delta)$  is a very sharp phase transition point for the density of the codes whose relative distance is greater than the given number  $\delta$ . Moreover, similar to Remark 3.2, for any point  $(\delta, r)$  of the rectangle  $(0, \delta_0) \times (0, 1)$  we can consider the asymptotic behavior of the probability  $\Pr(\Delta(C_r) > \delta)$  for the random code  $C_r$ . The rectangle can be divided into three areas by the curve  $r = \frac{1}{2}\text{GV}_q(\delta)$  (not the curve  $r = \text{GV}_q(\delta)$ !) and the curve for any upper bound as in Figure 4.1:

**Area I:** surrounded by the  $\delta$ -axis, the  $r$ -axis and the curve  $r = \frac{1}{2}\text{GV}_q(\delta)$  (the curve itself included), where “ $\Delta(C_r) > \delta$ ” is nearly a certain event.

**Area II:** surrounded by the curve  $r = \frac{1}{2}\text{GV}_q(\delta)$  and any upper bound (the dotted curve), where “ $\Delta(C_r) > \delta$ ” is a nearly impossible event.

**Area III:** above any upper bound (the dotted curve), where “ $\Delta(C_r) > \delta$ ” is an impossible event.

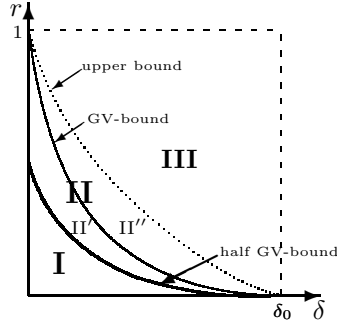


Figure 4.1 Three areas for the probability of the event “ $\Delta(C_r) > \delta$ ”

**Remark 4.4.** For an arbitrary random code  $C_r$  of rate  $r$  in  $F^n$ , Area II of Figure 4.1 can be further divided by the curve  $r = \text{GV}_q(\delta)$  into two subareas  $\text{II}'$  and  $\text{II}''$ , where the subarea  $\text{II}'$  is surrounded by the curve  $r = \frac{1}{2}\text{GV}_q(\delta)$  and the curve  $r = \text{GV}_q(\delta)$  (the curve itself included), and the subarea  $\text{II}''$  is the

remaining part of Area II; see Figure 4.1. The two subareas are distinct from each other in the following two points of view.

- For any  $(r, \delta)$  in the subarea II', though " $\Delta(C_r) > \delta$ " is nearly impossible, by the greedy algorithm one can construct codes of rate  $r$  and relative distance  $> \delta$ ; see Remark 2.1. On the other hand, the greedy algorithm does not work in the subarea II''.

- For the variable  $Y_{\mathbf{b}} = \hat{\mathcal{N}}_{C_r}^{g(\mathbf{b})}(\delta)$  defined in Eqn (4.8), from Remark 4.1(iii) one can, by taking  $t = 0$ , get that

$$\lim_{n \rightarrow \infty} \Pr(\hat{\mathcal{N}}_{C_r}^{g(\mathbf{b})}(\delta) = 0) = \begin{cases} 1, & r \leq r_\delta; \\ 0, & r > r_\delta; \end{cases}$$

where the range  $r \leq r_\delta = \text{GV}_q(\delta)$  covers Area I and the subarea II'. Comparing it with Corollary 4.2, one can see that the cumulative effect does change the threshold points.

**Remark 4.5.** Instead of the injective maps from  $F^k$  to  $F^n$  as in Eqn (4.5), one can consider the probability space whose sample space consists of all the maps  $\tilde{g} : F^k \rightarrow F^n$  and every sample is equipped with equiprobability; this probability space is just the so-called *random code ensemble* in Shannon's information theory, cf. [11, §6.1]. Then, for the random code  $\tilde{g}(F^k) \subseteq F^n$ , all the results of this section still hold.

## 5 Conclusion

For a random code  $C$  and  $\mathbf{c} \in C$ , the number of codewords with distance  $j$  from  $\mathbf{c}$ , denoted by  $\mathcal{N}_C^{\mathbf{c}}(j)$  and called the *distance enumerator with respect to  $\mathbf{c}$* , has been extensively studied with second moment analysis. In this paper we have introduced the *cumulative distance enumerators* of random codes, and investigated their thresholds by more sensitive second moment methods, including the Paley-Zygmund inequality. We have shown that the threshold phenomena of the relative distance and the GV-bound are just special extreme cases of the threshold phenomena of the cumulative distance enumerators.

For a random linear code  $L$  of length  $n$  and rate  $r$  over a finite field with  $q$  elements and a number  $\delta$  with  $0 < \delta < 1 - q^{-1}$ , we defined  $\hat{\mathcal{N}}_L^{\mathbf{0}}(\delta) = \sum_{j=1}^{\delta n} \mathcal{N}_L^{\mathbf{0}}(j)$  and its growth rate  $R(\hat{\mathcal{N}}_L^{\mathbf{0}}(\delta)) = \frac{\log_q(\hat{\mathcal{N}}_L^{\mathbf{0}}(\delta))}{n}$ . We have proved that the growth rate is asymptotically distributed at exactly one point  $r - r_\delta$ , where  $r_\delta$  denotes the GV-bound at  $\delta$ , see (3.10). Two consequences, which are more or less already known, were directly derived: the relative distance  $\Delta(L)$  of  $L$  is asymptotically distributed at the GV-distance  $\delta_r$ ; while the GV-bound  $r_\delta$  is a sharp threshold point for the density of linear codes whose relative distance is greater than  $\delta$ .

We consider an arbitrary random code  $C$  of length  $n$  and rate  $r$  over an alphabet with  $q$  elements in a similar manner. For a number  $\delta$  with  $0 < \delta < 1 - q^{-1}$ , we defined  $\hat{\mathcal{N}}_C(\delta) = \frac{1}{2} \sum_{\mathbf{c} \in C} \sum_{j=1}^{\delta n} \mathcal{N}_C^{\mathbf{c}}(j)$ , proved that the growth

rate  $R(\hat{\mathcal{N}}_C(\delta)) = \frac{\log_q(\hat{\mathcal{N}}_C(\delta))}{n}$  is asymptotically distributed at exactly one point  $2r - r_\delta$ , see (4.12); and deduced directly two consequences: the relative distance  $\Delta(C)$  is asymptotically distributed at the GV-distance  $\delta_{2r}$ ; half the GV-bound  $\frac{1}{2}r_\delta$  is a sharp threshold point for the density of codes whose relative distance is greater than  $\delta$ .

## Acknowledgements

The research of Y. Fan is supported by NSFC with grant number 11271005. The research of S. Ling and C. Xing is supported by the Singapore National Research Foundation Competitive Research Programme (grant number NRF-CRP2-2007-03). The research of H. W. Liu and J. Shen is supported by NSFC through grant number 11171370.

## References

- [1] A. Barg, G.D. Forney, “Random codes: minimum distances and error exponents”, *IEEE Trans. Inform. Theory*, vol.48, pp.2568-2573, 2002.
- [2] Y. Fan, J. Shen, “On the phase transitions of random  $k$ -constraint satisfaction problems”, *Artificial Intelligence*, vol.175, pp.914-927, 2011.
- [3] P. Gaborit, G. Zemor, “Asymptotic improvement of the Gilbert-Varshamov bound for linear codes”, *IEEE Trans. Inform. Theory*, vol.54, pp.3865-3872, 2008.
- [4] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [5] N.E. Gilbert, “A comparison of signalling alphabets”, *Bell Sys. Tech. Journal*, vol.31, pp.504-522, 1952.
- [6] R.W. Hamming, “Error detecting and error correcting codes”, *Bell Sys. Tech. Journal*, vol.29, pp.147-160, 1950.
- [7] W.C. Huffman, V. Pless, *Fundamentals of Error-Correcting Codes*, Cambridge University Press, 2003.
- [8] T. Jiang, A. Vardy, “Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes”, *IEEE Trans. Inform. Theory*, vol.50, pp.1655-1664, 2004.
- [9] A. Montanari, “The glassy phase of Gallager codes”, *Eur. Phys. J. B*, vol.23, pp.121-136, 2001.
- [10] N. Merhaev, “Relations between random coding exponents and the statistical physics of random codes”, *IEEE Trans. Inform. Theory*, vol.55, pp.83-92, 2009.

- [11] M. Mézard, A. Montanari, *Information, Physics and Computation*, Oxford University Press, 2009.
- [12] R.E.A.C. Paley and A. Zygmund, “A note on analytic functions in the unit circle”, *Proc. Camb. Phil. Soc.* vol.28, pp.266-272, 1932.  
Or, see [http://en.wikipedia.org/wiki/PaleyZygmund\\_inequality](http://en.wikipedia.org/wiki/PaleyZygmund_inequality)
- [13] J.N. Pierce, “Limit distribution of the minimum distance of random linear codes”, *IEEE Trans. Inform. Theory*, vol.13, pp.595-599, 1967.
- [14] C.E. Shannon, “A mathematical theory of communication”, *Bell Sys. Tech. Journal*, vol.27, pp.379-423, 623C655, 1948.
- [15] M.A. Tsfasman, S.G. Vladuts, T. Zink, “Modular curves, Shimura curves and Goppa codes, better than Varshamov-Gilbert bound”, *Math. Nachrichten*, vol.104, pp.13-28, 1982.
- [16] R.R. Varshamov, “Estimate of the number of signals in error-correcting codes (in Russian)”, *Dokl. Acad. Nauk*, vol.117, pp.739-741, 1957.
- [17] V. Vu, L. Wu, “Improving the Gilbert-Varshamov bound for  $q$ -ary codes”, *IEEE Trans. Inform. Theory*, vol.51, pp.3200-3208, 2005.
- [18] J. Wolfowitz, “The coding of messages subject to chance errors”, *Illinois J. Math.*, vol.1, pp.591-606, 1957.